

University of Groningen

Panel data models extended to spatial error autocorrelation or a spatially lagged dependent variable

Elhorst, J. Paul

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2001

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Elhorst, J. P. (2001). *Panel data models extended to spatial error autocorrelation or a spatially lagged dependent variable*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

PANEL DATA MODELS EXTENDED TO SPATIAL ERROR AUTOCORRELATION OR A SPATIALLY LAGGED DEPENDENT VARIABLE

J. Paul Elhorst¹

SOM-Theme C: Coordination and growth in economies

Abstract

This paper surveys panel data models extended to spatial error autocorrelation or a spatially lagged dependent variable. In particular, it focuses on the specification and estimation of four panel data models commonly used in applied research: the fixed effects model, the random effects model, the fixed coefficients model and the random coefficients model. This survey should prove useful for researchers in this area.

¹ Faculty of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands, tel. 31-50-3633893, fax. 31-50-3637337, e-mail: J.P.Elhorst@eco.rug.nl

1 INTRODUCTION

In spatial research, cross-sectional data often refer to observations made on a number of spatial units at a given time, time-series data to observations made over time on a given spatial unit, and panel data to observations made on a number of spatial units over time.² In recent years there has been a growing interest in the specification and estimation of econometric relationships based upon panel data. This interest can be explained by the fact that panel data offer researchers more possibilities than purely cross-sectional data or time-series data. According to Hsiao (1986) and Baltagi (1995), panel data give more informative data, more variability, less collinearity among the variables, more degrees of freedom, and more efficiency. Panel data also allow the specification of more complicated behavioral hypotheses, including effects that cannot be addressed using pure cross-sectional or time-series data.

Two problems may arise when panel data have a locational component. The first problem is spatial heterogeneity, which can be defined as parameters that are not homogeneous throughout the data set but vary with location.³ Parameter heterogeneity across spatial units has become a topical issue in the literature. Pesaran and Smith (1995) and Fotheringham et al. (1997)⁴, especially, advocate abandoning the fundamental assumption of homogeneous parameters underlying pooled models and relying upon average responses from individual regressions. As with cross-sectional data regression, the main problem of traditional panel data techniques is that they will only capture representative behavior; panel data regression with constant slopes takes an average across the spatial units, even when allowing for a variable intercept, and not the differing behavior of individual spatial units (Quah, 1996a, 1996b). A second reason why a relationship might exhibit spatial variation is that the model from which the relationship is being estimated is a gross misspecification of reality in that one or more relevant variables have been either omitted from the model or represented by an incorrect functional form and are making their presence felt through the parameter estimates.

It should be stressed that the rise of heterogeneous panel data estimators has also been resisted. First, because they seem to produce less plausible estimates than their pooled homogeneous counterparts (Baltagi and Griffin, 1997). Second, because estimating individual time series models — one for each spatial unit — leaves

²Other terms used to describe panel data are space-time data, longitudinal data, repeated-measures data and growth data.

³Coefficients may also vary over time, but that is not discussed in this paper.

⁴Fotheringham, Charlton and Brunsdon are known as advocates of geographically weighted regression.

undetected the co-movements across spatial units (Quah, 1996b). A final problem of heterogeneous panel data modeling is that it is only sensible to run separate regressions for each spatial unit when the number of observations on each spatial unit is large enough, while most panels do not meet this requirement.

The second problem that may arise when panel data have a locational component is that spatial dependence may exist between the observations at each point in time. The main reason that one observation associated with a location may depend on observations at other locations is that distance affects economic behavior. Each agent may change its economic decisions depending on the market conditions in the region of location compared to other regions and on the distance to these regions. These notions have been formulated in regional science theory, which relies on notions of spatial interaction and diffusion effects, hierarchies of place and spatial spillovers. To model spatial dependence between observations, the model may take the form of a spatial autoregressive process in the error term or in the variable to explain. The first is known as the *spatial error* case and the second as the *spatial lag* case.⁵

This paper surveys panel data models extended to either spatial error autocorrelation or a spatially lagged dependent variable. The reason for this paper is that these kinds of panel data models are not very well documented in the literature. Only Anselin (1988), in his seminal textbook on spatial econometrics, discusses some panel data models including spatial effects.⁶ Besides, there are also some empirical studies allowing varying coefficients across spatial units together with spatial effects, but these studies only give the most obvious or easily understood features of the method of estimation, and not those that require more effort to apply these methods in practice (e.g. Case, 1991; Kelejian and Robinson, 1997; Buettner, 1999; among others).

The spatial econometric literature has shown that OLS estimation in models with spatial effects is inappropriate. In the case of spatial error autocorrelation, the OLS estimator of the response parameters, while unbiased, loses its property of efficiency. In the case of a spatially lagged dependent variable, the OLS estimator of the response parameters not only loses its property of unbiasedness but also its consistency. The

⁵for the introduction of these terms, see Anselin and Hudak (1992).

⁶On pp.150-156 and pp.164-166 Anselin presents the random effects model including spatial error autocorrelation, and on pp.137-150 and pp.157-163 the seemingly unrelated regressions model including a spatial lagged dependent variable and spatial error autocorrelation. The difference with the seemingly unrelated regressions model presented in this paper is that, while Anselin allows the coefficients to vary across time, we allow them to vary across space. Other models presented in this paper that are briefly mentioned in Anselin are the random coefficients model including spatial error autocorrelation (pp.129-131), and the simultaneous equations model with one separate equation for every spatial unit including spatial lagged dependent variables (p.156).

latter might be thought of as the minimum requirement for a useful estimator. The most commonly suggested method to overcome these problems is to estimate the model by maximum likelihood (see Anselin, 1988; Anselin and Hudak, 1992). For this reason, we will deal with maximum likelihood estimation in principle, unless this approach is too complicated or not applicable.

The plan of this paper is as follows: First, we introduce the four most commonly used panel data models for estimating relationships that are linear in the parameters and in which the regressand is a continuous variable. Then we extend these models to spatial error autocorrelation or a spatially lagged dependent variable and explain how these models can be estimated. Finally, we present our conclusions.

2 A TAXONOMY OF TYPICAL PANEL DATA MODELS

The starting point of many econometric analyses in spatial research is a linear model between a dependent variable Y and a set of K independent variables X

$$Y_{it} = \beta_1 X_{1it} + \beta_2 X_{2it} + \dots + \beta_K X_{Kit} + \varepsilon_{it} = \beta' X_{it} + \varepsilon_{it}, \quad (1)$$

where $i=1, \dots, N$ refers to a spatial unit, $t=1, \dots, T$ refers to a given time period, β_1, \dots, β_K are fixed but unknown parameters, ε_{it} are error terms independently and identically distributed (i.i.d.) for all i and t with zero mean and variance σ^2 . Usually, one of the independent variables is fixed at unity, say X_{1it} for all i and t , and its parameter β_1 is called the intercept of the model.

The main objection to this model is that it does not control for spatial heterogeneity. Spatial units might differ in their background variables, mostly space-specific time-invariant variables that do affect the dependent variable of the analysis but are difficult to measure or hard to obtain. To fail to account for these variables runs the risk of obtaining biased results. One remedy is to introduce a variable intercept μ_i representing the effect of the omitted variables that are peculiar to each spatial unit under study

$$Y_{it} = \beta' X_{it} + \mu_i + \varepsilon_{it}. \quad (2)$$

Conditional upon the specification of this variable intercept, the regression equation can be estimated as a fixed effects or random effects model. In the fixed effects model, a dummy variable is introduced for each spatial unit as a measure of the variable intercept. In the random effects model, the variable intercept is treated as a random variable i.i.d. with zero mean and variance σ_μ^2 . Furthermore, it is assumed that the μ_i

and ε_{it} random variables are independent of each other. Another difference between these two models is that for short panels, where T is fixed and $N \rightarrow \infty$, the fixed effects model suffers from a loss of degrees of freedom in that only the slope parameters can be estimated consistently; the coefficients of the dummy variables cannot be estimated consistently since the number of these coefficients increases as N increases. In the random effects model, this loss of degrees of freedom is avoided.

Although the variable intercept model accommodates for the effect of spatial heterogeneity to a certain degree, the problem remains as to whether the data in such a model are pooled correctly. When spatial heterogeneity cannot be captured completely by the variable intercept, a natural generalization would be to let the slope parameters of the regressors vary as well. Just like the variable intercept, the slope parameters may be considered as fixed or as randomly distributed between spatial units. If the parameters are fixed and different for different spatial units, each spatial unit is treated separately. Let $Y_i = X_i \beta_i + \varepsilon_i$ be the i th equation in the set of N equations, where the observations are stacked as an equation for each spatial unit over time. The only way to relate these N separate regressions is to assume correlation between the error terms in different equations, known as contemporaneous error correlation. Such a specification is reasonable in case the error terms for different spatial units, at a given point in time, are likely to reflect some common unmeasurable or omitted factors. In full-sample notation, the set of N equations may be written as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \cdot & 0 \\ 0 & X_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad E(\varepsilon_i) = 0, E(\varepsilon_i \varepsilon_j) = \sigma_{ij} I_T, \quad i, j = 1, \dots, N. \quad (3)$$

This model is known as a seemingly unrelated regressions (SUR) model.

A disadvantage of a model with different parameters for different spatial units is the large number of parameters to be estimated: $N \times K$ different β parameters and $\frac{1}{2}N(N+1)$ different σ parameters. This model is therefore only of value when T is large and N is small. If T were fixed and $N \rightarrow \infty$, as is typical in short panels, this model would not be of value in that the response parameters could not be estimated consistently. This is because the number of these parameters increases as N increases.

If the parameters are treated as outcomes of random experiments between spatial units, we can pool the data into one model to estimate the unknown parameters. This is known as a Swamy type random coefficients model (Swamy, 1970)

$$Y_{it} = \beta_i' X_{it} + \varepsilon_{it}, \quad E(\varepsilon_{it}) = 0, \quad E(\varepsilon_{it} \varepsilon_{ts}) = \sigma_i^2 \quad (t, s = 1, \dots, N), \quad i = 1, \dots, N, \quad (4a)$$

$$\beta_i = \beta + v_i, \quad i = 1, \dots, N, \quad (4b)$$

where the β_i applying to a particular spatial unit is the outcome of a random process with common-mean-coefficient vector $E(v_i)=0$, $E(v_i v_i')=V$, and V is a symmetric $K \times K$ matrix. Furthermore, it is assumed that $E(v_i v_j')=0$ for $i \neq j$ and that the ε_{it} and v_i random variables are independent of each other. By treating the differences in the response coefficients across spatial units as outcomes of random experiments instead of fixed coefficients, the number of response coefficients no longer grows with the number of spatial units. This substantially reduces the number of parameters to be estimated and improves the efficiency of the estimates due to the availability of many more degrees of freedom.

3 PANEL DATA MODELS EXTENDED TO SPATIAL ERROR OR LAG

It proves helpful to introduce the following notation: Let W denote a $N \times N$ spatial weight matrix describing the spatial arrangement of the spatial units and w_{ij} the (i,j) th element of W ($i,j=1,\dots,N$). It is assumed that W is a matrix of known constants, that all diagonal elements of W are zero ($w_{ii}=0$, $i=1,\dots,N$), and that the characteristic roots of W ($\omega_1, \dots, \omega_N$) are known. The first assumption excludes the possibility that the spatial weight matrix is parametric; the second implies that no spatial unit can be viewed as its own neighbor; and the third presupposes that the characteristic roots of W can be computed accurately based on computing technology typically available to empirical researchers⁷ and is needed to ensure that the log-likelihood function of the models below can be computed.

Fixed effects model

If we stack the observations as one equation for each cross-section at one point in time, the fixed effects model extended to spatial error autocorrelation can be specified as

$$Y_t = X_t \beta + \mu + \varphi_t, \quad \varphi_t = \delta W \varphi_t + \varepsilon_t, \quad E(\varepsilon_t) = 0, \quad E(\varepsilon_t \varepsilon_t') = \sigma^2 I_N, \quad (5)$$

and to a spatially lagged dependent variable as

⁷Kelejian and Prucha (1999) have pointed out that this might be problematic even for moderate sample sizes ($N=400$).

$$Y_t = \delta WY_t + X_t\beta + \mu + \varepsilon_t, \quad E(\varepsilon_t) = 0, \quad E(\varepsilon_t \varepsilon_t') = \sigma^2 I_N, \quad (6)$$

where $\mu = (\mu_1, \dots, \mu_N)'$. Note that in the spatial error case the properties of the error structure have been changed, and that in the spatial lag case the number of explanatory variables has increased. In both cases δ is called the spatial autoregressive coefficient.

The standard method of estimating the fixed effects model is to eliminate the intercept and μ_i from the regression equation by demeaning the Y and X variables⁸, estimate the resulting demeaned equation by OLS, and then recover the intercept β_1 and μ_i (Baltagi, 1995, pp.10-13). It should be noted that only $(\beta_1 + \mu_i)$ are estimable, and not β_1 and μ_i separately, unless a restriction such as $\sum \mu_i = 0$ is imposed.

Instead of estimating the demeaned equation by OLS, it can also be estimated by ML. The only difference is that ML estimators do not make corrections for degrees of freedom. The log-likelihood function corresponding to the demeaned equation extended to spatial error autocorrelation is

$$-\frac{NT}{2} \ln(2\pi\sigma^2) + T \sum_{i=1}^N \ln(1 - \delta\omega_i) - \frac{1}{2\sigma^2} \sum_{t=1}^T e_t' e_t, \quad e_t = (I - \delta W)[Y_t - \bar{Y} - (X_t - \bar{X})\beta], \quad (7)$$

and to a spatially lagged dependent variable is

$$-\frac{NT}{2} \ln(2\pi\sigma^2) + T \sum_{i=1}^N \ln(1 - \delta\omega_i) - \frac{1}{2\sigma^2} \sum_{t=1}^T e_t' e_t, \quad e_t = (I - \delta W)(Y_t - \bar{Y}) - (X_t - \bar{X})\beta. \quad (8)$$

These two log-likelihood functions can be maximized by standard techniques developed by Anselin (1988, pp.181-182) and Anselin and Hudak (1995).

Random effects model

To describe the estimation procedure of the random effects model, it proves helpful to separate the intercept β_1 from the other right-hand side variables. This implies that X_t and β in this particular case denote all the regressors and their response parameters except the intercept. Just as in the fixed effects model, *we stack the observations as one equation for each cross-section at one point in time.*

Starting with the log-likelihood function of the random effects model given in

⁸Each variable for every spatial unit is taken as the deviation of its average over time:

$$Y_{it} - \bar{Y}_i \quad \text{with} \quad \bar{Y}_i = \frac{1}{T} \sum_{t=1}^T Y_{it}, \quad i = 1, \dots, N; \text{ and the same for the X variables.}$$

Breusch (1987) and Baltagi (1995, pp.18-19), the log-likelihood function of the random effects model extended to spatial error autocorrelation is

$$-\frac{NT}{2}\ln(2\pi\sigma^2) + T\sum_{i=1}^N \ln(1 - \delta\omega_i) + \frac{N}{2}\ln\theta^2 - \frac{1}{2\sigma^2}\sum_{t=1}^T e_t'e_t,$$

with $e_t = (I - \delta W)[Y_t - (1 - \theta)\bar{Y} - \theta\beta_1 - (X_t - (1 - \theta)\bar{X})\beta]$, (9)

and to a spatially lagged dependent variable is

$$-\frac{NT}{2}\ln(2\pi\sigma^2) + T\sum_{i=1}^N \ln(1 - \delta\omega_i) + \frac{N}{2}\ln\theta^2 - \frac{1}{2\sigma^2}\sum_{t=1}^T e_t'e_t,$$

with $e_t = (I - \delta W)(Y_t - (1 - \theta)\bar{Y}) - \theta\beta_1 - (X_t - (1 - \theta)\bar{X})\beta$. (10)

The parameter θ^2 measures the weight that will be attached to the variation between the spatial units.⁹ If this weight tends to zero, the random effects model reduces to the fixed effects model, a model that only utilizes the variation within the spatial units over time in forming the parameter estimates of β . If the weight tends to unity, the random effects model reduces to the standard model in (1) estimated by OLS.

Both β_1 and σ^2 can be removed by concentrating the log-likelihood functions, i.e., by substituting the unique maximizing values of β_1 and σ^2 , given arbitrary values of β , δ and θ^2 . Estimates of β_1 and σ^2 can be recovered later by

$$\beta_1 = \tilde{Y} - \beta_2\tilde{X}_2 - \dots - \beta_K\tilde{X}_K, \text{ with } \tilde{Y} = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T Y_{it} \text{ and } \tilde{X}_k = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T X_{kit}, k = 2, \dots, K,$$
(11a)

$$\sigma^2 = \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T e_{it}^2 = \frac{1}{T}\sum_{t=1}^T e_t'e_t.$$
(11b)

The concentrated log-likelihood function for the spatial error case is then

⁹Instead of σ_μ^2 the parameter θ^2 is estimated, which is possible as there is a one-to-one correspondence between θ^2 and σ_μ^2 : $\theta^2 = \sigma^2 / (\sigma^2 + T\sigma_\mu^2)$. θ is defined as $\theta = \sqrt{\theta^2}$.

$$-\frac{NT}{2}\ln(2\pi) - \frac{T}{2} + \frac{NT}{2}\ln T + T \sum_{i=1}^N \ln(1 - \delta\omega_i) + \frac{N}{2}\ln \theta^2 - \frac{NT}{2}\ln\left(\sum_{t=1}^T d_t' d_t\right),$$

with $d_t = (I - \delta W)[Y_t - (1 - \theta)\bar{Y} - \theta\tilde{Y} - (X_t - (1 - \theta)\bar{X} - \theta\tilde{X})\beta]$, (12)

and for the spatial lag case is

$$-\frac{NT}{2}\ln(2\pi) - \frac{T}{2} + \frac{NT}{2}\ln T + T \sum_{i=1}^N \ln(1 - \delta\omega_i) + \frac{N}{2}\ln \theta^2 - \frac{NT}{2}\ln\left(\sum_{t=1}^T d_t' d_t\right),$$

with $d_t = (I - \delta W)(Y_t - (1 - \theta)\bar{Y} - \theta\tilde{Y}) - (X_t - (1 - \theta)\bar{X} - \theta\tilde{X})\beta$. (13)

In both cases one can iterate between β on the one hand and θ^2 and δ on the other. The estimator of β , given θ^2 and δ , is a GLS estimator; this estimator can be obtained by OLS regression of the transformed variable Y on the transformed variables X . Conversely, the estimators of θ^2 and δ , given β , must be solved by numerical methods as they cannot be solved algebraically. As the basic regression model extended to spatial error autocorrelation or to a spatially lagged dependent variable has one maximum for δ ¹⁰, so does the random effects model. By contrast, Maddala (1971) has pointed out that the log-likelihood function of the random effects model without spatial effects may have two maxima for $0 < \theta^2 \leq 1$. To guard against the possibility of a local maximum, Breusch (1987) has pointed out that one should start iterating with the "within estimator" ($\theta^2 = 0$) and the "between estimator" ($\theta^2 \rightarrow \infty$). If these two sequences converge to the same maximum, then this is the global maximum. Finally, it should be noted that estimating the random effects model by ML is not standard in that this application is not available in most commercial econometric software packages.

Heterogeneous model with fixed coefficients

Both in the spatial error case and in the spatial lag case we assume that the observations are stacked as one equation for each cross-section at one point in time.

Spatial error. The SUR model in (3) with one equation for every spatial unit and with contemporaneous error correlation does not have to be changed to cope with the spatial error case since the set of σ_{ij} ($i, j = 1, \dots, N$) already reflects the interactions between the spatial units. In the literature this is contemplated as an advantage as no prior assumptions are required about the nature of interactions over space (White and Hewings, 1982). The explanation is that the specification of a particular spatial weight matrix does not alter the estimates of the response parameters β ; the estimate of each σ_{ij} immediately adapts itself to the value of w_{ij} by which it would be multiplied. As the SUR model is discussed in almost every econometric textbook and is available in

¹⁰Equations (5) and (6) but without the variable intercept μ (see Anselin and Hudak, 1992).

almost every commercial econometric software package, it hardly requires any further explanation.

The standard method of obtaining the maximum likelihood estimates of the parameters in a SUR model is by iterating the feasible GLS procedure. In every iteration the residuals of the separate regressions are used to update the elements of the covariance matrix

$$\sigma_{ij} = e_i' e_j / T, \quad i, j = 1, \dots, N, \quad (14)$$

until convergence. It is to be noted that the estimates of β_i ($i=1, \dots, N$) and σ_{ij} ($i, j=1, \dots, N$) that are obtained by iterating the feasible GLS procedure are equivalent to those that would be obtained by maximizing the log-likelihood function of the model, assuming that there are no restrictions on the response parameters β across or within the equations.

One alternative form of the SUR model, which re-establishes the spatial weight matrix, is obtained by imposing the restrictions $\sigma_{ij} = \delta_i w_{ij}$ for $j \neq i$. These restrictions are reasonable if one has prior information about the nature of interactions over space. Furthermore, they reduce the number of parameters to be estimated, which is of help especially when N is large (e.g., $N > 20$). On imposing the restrictions $\sigma_{ij} = \delta_i w_{ij}$ for $j \neq i$, the elements of the covariance matrix must be updated in each iteration by

$$\sigma_{ii} = \frac{e_i' e_i}{T}, \quad \delta_i = \frac{\sum_{j=1, j \neq i}^N w_{ij} e_i' e_j}{T \sum_{j=1, j \neq i}^N w_{ij}}, \quad i = 1, \dots, N. \quad (15)$$

These restrictions can easily be implemented.

Spatial lag. The set of N equations in a heterogeneous model with fixed coefficients and spatially lagged dependent variables can be expressed as

$$\begin{bmatrix} Y_1 & Y_2 & \dots & Y_N \end{bmatrix} \begin{bmatrix} 1 & -\delta_{21} & \dots & -\delta_{N1} \\ -\delta_{12} & 1 & \dots & -\delta_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ -\delta_{1N} & -\delta_{2N} & \dots & 1 \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_N \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}, \quad (16)$$

or

$$Y\Gamma = XB + \varepsilon, \quad E(\varepsilon) = 0, \quad E(\varepsilon\varepsilon') = \Sigma \otimes I_T \quad \text{with } \Sigma = \sigma_{ii} I_N \quad (i = 1, \dots, N). \quad (17)$$

Each equation can also be written as

$$Y_i = [Y_1 \dots Y_{i-1} Y_{i+1} \dots Y_N X_i] \begin{bmatrix} \delta_{i1} \\ \vdots \\ \delta_{ii-1} \\ \delta_{ii+1} \\ \vdots \\ \delta_{iN} \\ \beta_i \end{bmatrix} + \varepsilon_i = Z_i \eta_i + \varepsilon_i, \quad i = 1, \dots, N. \quad (18)$$

Note (i) that the δ s, just as the β s, are assumed to be different for different spatial units, (ii) that no prior assumptions are required about the nature of interactions over space, just as in the SUR model, and (iii) that this model is a simplified version of a linear simultaneous equations model in that the assumption of contemporaneous error correlation has been dropped. This assumption has been dropped to discriminate between the spatial error case and the spatial lag case.

The log-likelihood function and the first-order maximizing conditions of a linear simultaneous equations model are given in Hausman (1975, 1983). Dropping the assumption of contemporaneous error correlation, the FIML estimator of each single η_i is

$$\eta_i = \begin{bmatrix} \delta_i \\ \beta_i \end{bmatrix} = (\hat{Z}_i' Z_i)^{-1} \hat{Z}_i' Y_i, \quad i = 1, \dots, N, \quad (19a)$$

$$\text{where } \hat{Z}_i = [(XB\Gamma^{-1})_i \ X_i], \text{ while } \sigma_{ii} = \frac{(Y_i - Z_i \eta_i)'(Y_i - Z_i \eta_i)}{T}, \quad i = 1, \dots, N. \quad (19b)$$

The matrix of estimated values of Z_i consists of $N-1+K$ columns: $N-1$ columns with respect to those spatially lagged dependent variables that explain Y_i ¹¹ and K columns with respect to those independent variables that explain Y_i . Note that the estimated values of Z_i can also be seen as instrumental variables (Hausman, 1975, 1983). As the estimated values of Z_i at the right-hand side of (19) depend on η , (19) defines no closed form solution for η . One might attempt to solve η by the Jacobi iteration method¹²,

¹¹The matrix $XB\Gamma^{-1}$ consists of N columns. In case Y_j ($j=1, \dots, N$) is an explanatory variable of Y_i ($i=1, \dots, N$), the j th column of $XB\Gamma^{-1}$ is part of the matrix of estimated values of Z_i .

¹²We require a solution $\eta=f(\eta)$. The Jacobi iteration method iterates according to $\eta^{h+1}=f(\eta^h)$.

though a problem is that this method cannot be expected to converge in general. Furthermore, this procedure is available in only a limited number of commercial econometric software packages.¹³

One alternative form of this model, which re-establishes the spatial weight matrix, is obtained by also imposing the restrictions $\delta_{ij} = \delta_i w_{ij}$ for $j \neq i$. Just as in the spatial error case, these restrictions are reasonable if one has prior information about the nature of interactions over space. Furthermore, they reduce the number of parameters to be estimated, which is of help especially when N is large (e.g., $N > 20$). Under these circumstances we get

$$\Gamma = \begin{bmatrix} 1 & -\delta_2 w_{21} & \cdot & -\delta_N w_{N1} \\ -\delta_1 w_{21} & 1 & \cdot & -\delta_N w_{N2} \\ \cdot & \cdot & \cdot & \cdot \\ -\delta_1 w_{N1} & -\delta_2 w_{2N} & \cdot & 1 \end{bmatrix} \text{ and } \hat{Z}_i = \sum_{j=1}^N w_{ij} [XB\Gamma^{-1}]^j, i = 1, \dots, N, \quad (20)$$

where $[XB\Gamma^{-1}]^j$ denotes the j th column of the matrix $XB\Gamma^{-1}$. Although these restrictions simplify the estimation procedure, it follows that the Jacobi iteration method is still needed. Because a heterogeneous model with fixed parameters has different spatial autoregressive coefficients δ for different spatial units, it also follows that the Jacobian term, in contrast to both the fixed effects and random effects models that have one spatial autoregressive coefficient δ for all spatial units, cannot be expressed in function of the characteristic roots of the spatial weight matrix. It is this difference that complicates the numerical determination of the FIML estimator. As an alternative, one might also use 2SLS since this estimator has the same asymptotic distribution as the FIML estimator. The benefit of the 2SLS estimator is that it is far easier to compute. The cost is a loss in asymptotic efficiency, as it does not take account of the restrictions on the coefficients within the matrices B and Γ .

Heterogeneous model with random coefficients

Spatial error. If we stack the observations as an equation for each spatial unit over time, the random coefficient model with spatial error can be specified as

$$Y_i = X_i \beta_i + \varepsilon_i, \quad E(\varepsilon_i) = 0, \quad E(\varepsilon_i \varepsilon_j) = \sigma_{ij} I_T \quad (j = 1, \dots, N), \quad i = 1, \dots, N, \quad (21a)$$

$$\beta_i = \beta + v_i, \quad i = 1, \dots, N, \quad (21b)$$

where the β_i applying to a particular spatial unit is the outcome of a random process

¹³Examples are TSP and PC-Give (see Greene, 1997).

with common-mean-coefficient vector $E(v_i)=0$, $E(v_i v_i')=V$, and V is a symmetric $K \times K$ matrix. Furthermore, it is assumed that $E(v_i v_j')=0$ for $i \neq j$ and that the ε_{it} and v_i random variables are independent of each other.

Just as in the heterogeneous model with fixed coefficients, no prior assumptions are required about the nature of interactions over space. Note that compared to the Swamy random coefficient model, we assume not only that $E(\varepsilon_{it} \varepsilon_{is}) = \sigma_{ii}^{14}$ ($t, s=1, \dots, T$) for $i=1, \dots, N$, but also that $E(\varepsilon_{it} \varepsilon_{js}) = \sigma_{ij}$ ($t, s=1, \dots, T$) for $i, j=1, \dots, N$. In this model, the random vector $Y \equiv (Y_1', \dots, Y_N')$ can be seen to be distributed with mean $X\beta$, where $X \equiv (X_1', \dots, X_N')$, and covariance matrix

$$\Sigma = \begin{bmatrix} X_1' V X_1 + \sigma_{11} I_T & \sigma_{12} I_T & \cdot & \sigma_{1N} I_T \\ \sigma_{12} I_T & X_2' V X_2 + \sigma_{22} I_T & \cdot & \sigma_{2N} I_T \\ \cdot & \cdot & \cdot & \cdot \\ \sigma_{1N} I_T & \sigma_{N2} I_T & \cdot & X_N' V X_N + \sigma_{NN} I_T \end{bmatrix} = D(I_N \otimes V)D' + (\Sigma_\sigma \otimes I_T), \quad (22)$$

where D is a block-diagonal matrix of $NT \times NK$, $D = \text{diag}[X_1, \dots, X_N]$, and Σ_σ is a $N \times N$ matrix, $\Sigma_\sigma = \{\sigma_{ij}\}$ ($i, j=1, \dots, N$). The ML and the GLS estimator of β are known to be equivalent (Lindstrom and Bates, 1988) and equal to

$$\hat{\beta} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} Y, \quad (23)$$

though the problem is that Σ contains unknown parameters $\Sigma = \Sigma(\Sigma_\sigma, V)$ that must also be estimated. There are two ways to proceed. A feasible GLS estimator of β may be constructed based on a consistent estimate of Σ_σ and V . To obtain this estimator, the following steps must be carried out: First, estimate the model assuming that all the response parameters are fixed and different for different spatial units. We use the abbreviation FC to refer to these estimates. The model in view is the SUR model without restrictions on the covariance matrix (equations (3) and (14)), or the SUR model with the restrictions $\sigma_{ij} = \delta_i w_{ij}$ for $j \neq i$ on the covariance matrix (equations (3) and (15)). This results in estimates for β_i^{FC} and σ_{ij}^{FC} ($i, j=1, \dots, N$) or β_i^{FC} , σ_{ii}^{FC} and δ_i^{FC} ($i=1, \dots, N$). Second, estimate V by (see Swamy, 1974)

¹⁴In this case we decided to use σ_{ii} instead of σ_i^2 as in equation (4a).

$$V = \frac{1}{N-1}S - \frac{1}{N} \sum_{i=1}^N \hat{\sigma}_{ii}^{FC} (X_i' X_i)^{-1} + \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1, j \neq i}^N s_{ij} (X_i' X_i)^{-1} X_i' X_j (X_j' X_j)^{-1},$$

where $S = \sum_{i=1}^N (\hat{\beta}_i^{FC} - \frac{1}{N} \sum_{i=1}^N \hat{\beta}_i^{FC}) (\hat{\beta}_i^{FC} - \frac{1}{N} \sum_{i=1}^N \hat{\beta}_i^{FC})'$ and $s_{ij}^{FC} = \hat{\sigma}_{ij}^{FC}$ or $\hat{\delta}_{ij}^{FC} w_{ij}$. (24)

The estimator of V , though unbiased, may not be positive definite. To ensure the positive definiteness of the estimated matrix, one might also use the consistent estimator $V=1/(N-1) \times S$ (for details, see Swamy, 1970). Finally, estimate the common-mean coefficient vector β by GLS according to equation (23). One problem of the final step is that it requires a matrix inversion of order $N \times T$. As an alternative, the inverse of Σ can also be computed by the expression

$$\Sigma^{-1} = (\Sigma_{\sigma}^{-1} \otimes I_T) - (\Sigma_{\sigma}^{-1} \otimes I_T) D [D' (\Sigma_{\sigma}^{-1} \otimes I_T) D + I_N \otimes V^{-1}]^{-1} D' (\Sigma_{\sigma}^{-1} \otimes I_T), \quad (25)$$

which requires the inversion of three matrices, one of order K (V), one of order N (Σ_{σ}) and one of order $N \times K$ (the matrix between square brackets). In the case that T is large and/or $K \ll T$, this alternative computation is to be preferred. Nevertheless, the inversion of a matrix of order $N \times K$ in some commercial econometric software packages still might pose computational difficulties.

Despite its mathematical equivalence, the feasible GLS estimator of β does not coincide with the ML estimator of β . This is because the feasible GLS estimator of β is based on a consistent but not on the ML estimate of Σ_{σ} and V . Following the statistical literature, ML estimation of β , Σ_{σ} and V , though possible, seems laborious. There are three reasons for this. First, Σ_{σ} and V cannot be solved algebraically from the first-order maximizing conditions of the log-likelihood function. Consequently, Σ_{σ} and V must be solved by numerical methods. Second, a common estimation problem is associated with the restrictions on the parameters of the covariance matrix. A variance estimate should be nonnegative, and a covariance matrix estimate should be nonnegative definite. Also, it must be allowed that an estimate takes values on the boundary. Thus, a variance estimate may be zero, and a covariance matrix estimate may be a nonnegative definite matrix of any rank. In fact, such boundary cases provide useful exploratory information during the model building. It is desirable that numerical algorithms for ML estimators can successfully produce the defined estimates for all possible samples including those where the maximum is attained on the boundary. However, this parameter space problem often causes difficulties for the existing ML algorithms (Shin and Amemiya, 1997, p.190). Third, although some studies assert to have developed efficient and effective algorithms for the likelihood-based estimation of the parameters, they generally assume that $E(\varepsilon_i \varepsilon_i') = \sigma^2 I_T$ and $E(\varepsilon_i \varepsilon_j') = 0$ for $i, j = 1, \dots, N$ and $i \neq j$ instead of $E(\varepsilon_i \varepsilon_j') = \sigma_{ij}$ (Jenrich and Schluchter, 1986; Lindstrom and Bates, 1988, p.1014, left

column; Longford, 1993; Goldstein, 1995; Shin and Amemiya, 1997, p.189). This naturally simplifies the parameter space problem. It is therefore not certain whether these algorithms work for this more general case.

Spatial lag. A full random parameter model with spatially lagged dependent variables does in fact not exist. The main reason for this is that the assumption of a random element in the coefficients of lagged dependent variables raises intractable difficulties at the level of identification and estimation (Kelejian, 1974; Balestra and Negassi, 1992; Hsiao, 1996). Instead a mixed model may be used, containing fixed coefficients with respect to the spatial dependent variables and random coefficients with respect to the exogenous variables. This model reads as

$$Y_{it} = \delta_{li} Y_{lt} + \dots + \delta_{ii-1} Y_{i-1t} + \delta_{ii+1} Y_{i+1t} + \dots + \delta_{Ni} Y_{Nt} + \beta_i' X_{it} + \varepsilon_{it}, \quad (26a)$$

$$\beta_i = \beta + v_i, \quad (26b)$$

where $E(\varepsilon_{it})=0$, $E(\varepsilon_{it}\varepsilon_{is})=\sigma_i^2$ ($t,s=1,\dots,T$; $i=1,\dots,N$), and where the β_i applying to a particular spatial unit is the outcome of a random process with common-mean-coefficient vector $E(v_i)=0$, $E(v_i v_i')=V$, and V is a symmetric $K \times K$ matrix. Furthermore, it is assumed that $E(v_i v_j')=0$ for $i \neq j$ and that the ε_{it} and v_i random variables are independent of each other.

A problem making this model still quite unusual is the number of observations needed for its estimation. The minimum number of observations on each spatial unit amounts to $N+K$, as the number of regressors is $N-1+K$. Most panels do not meet this requirement, even if N is small. Provided that information is available about the nature of interactions over space, we therefore impose the restrictions $\delta_{ij}=\delta_i w_{ij}$ for $j \neq i$, to get

$$Y_{it} = \delta_i \sum_{j=1}^N w_{ij} Y_{jt} + \beta_i' X_{it} + \varepsilon_{it} \equiv \delta_i Y_i(w) + \beta_i' X_{it} \equiv \eta_i' Z_{it} + \varepsilon_{it}. \quad (26a')$$

In this case the minimum number of observations needed on each spatial unit reduces to $K+1$, which is independent of N .

Stacking the observation by time for each spatial unit, the full model may be expressed as

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} = \begin{bmatrix} Y_1(w) & 0 & \cdot & 0 \\ 0 & Y_2(w) & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & Y_N(w) \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \cdot \\ \delta_N \end{bmatrix} + \begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ X_N \end{bmatrix} \beta + \begin{bmatrix} X_1 & 0 & \cdot & 0 \\ 0 & X_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & X_N \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \cdot \\ v_N \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \varepsilon_N \end{bmatrix} \equiv$$

$$\equiv \text{diag}[Y_1(w), \dots, Y_N(w)] \times \delta + X\beta + \text{diag}[X_1, \dots, X_N] \times v + \varepsilon. \quad (27a)$$

The covariance matrix of the composite disturbance term $\text{diag}[X_1, \dots, X_N] \times v + \varepsilon$ is block-diagonal with the i^{th} diagonal block given by

$$\Phi_i = X_i V X_i' + \sigma_i^2 I_T. \quad (27b)$$

Just as in the spatial error case, there are two ways to proceed. A feasible GLS estimator of δ and β may be constructed, extended to instrumental variables and based on a consistent estimate of $\sigma_1^2, \dots, \sigma_N^2$ and V . Alternatively, $\delta, \beta, \sigma_1^2, \dots, \sigma_N^2$ and V may be estimated by ML.¹⁵ We suggest the following feasible GLS analog instrumental variables estimator taken from Bowden and Turkington (1984, Ch.3).¹⁶

Let X_i denote the $T \times K$ matrix of the exogenous variables in the i th equation, Z_i the $T \times (1+K)$ matrix of the spatially lagged dependent and exogenous variables in the i th equation, and X denote the $T \times K_{\text{ALL}}$ matrix of all the exogenous variables in the full model. If each spatial unit has K random regressors, among which the intercept, K_{ALL} is equal to $1+N(K-1)$. Consequently, the inversion of the matrix $X'X$ of order $K_{\text{ALL}} \times K_{\text{ALL}}$ might be a problem when N and K are large.

First, estimate the model assuming that all coefficients are fixed. We use the abbreviation FC to refer to these estimates. The model in view might be the SUR model extended to spatially lagged dependent variables described above (equations (16)-(18) combined with (20)), but in this case we stick to instrumental variables estimators. This results in estimates for β_i^{FC} and $\sigma_i^{2,\text{FC}}$ ($i, j=1, \dots, N$)

$$\hat{\eta}_i^{\text{FC}} = \begin{bmatrix} \hat{\delta}_i^{\text{FC}} \\ \hat{\beta}_i^{\text{FC}} \end{bmatrix} = [Z_i' X (X'X)^{-1} X' Z_i]^{-1} Z_i' X (X'X)^{-1} X Y_i, \quad (28a)$$

$$\hat{\sigma}_i^{2,\text{FC}} = \frac{(Y_i - Z_i \hat{\eta}_i^{\text{FC}})' (Y_i - Z_i \hat{\eta}_i^{\text{FC}})}{T - K}. \quad (28b)$$

¹⁵We have found one application of this model in the literature (Sampson et al., 1999), but this paper omits a description of the estimation procedure used.

¹⁶Bowden and Turkington start with the regression equation $Y = X\beta + u$, where $E(uu') = \Omega$. Part of the X variables are endogenous. Let Z denote the instrumental variables. Then the GLS analog instrumental variables estimator is

$$\hat{\beta} = (X^p' \Omega^{-1} X^p)^{-1} X^p' \Omega^{-1} Y \quad \text{with} \quad X^p = Z(Z' \Omega^{-1} Z)^{-1} Z' \Omega^{-1} X.$$

Second, estimate V by (see Balestra and Negassi, 1992; Hsiao and Tahmiscioglu, 1997)

$$\hat{V} = \frac{1}{N-1} \sum_{i=1}^N (\hat{\beta}_i^{FC} - \frac{1}{N} \sum_{i=1}^N \hat{\beta}_i^{FC}) (\hat{\beta}_i^{FC} - \frac{1}{N} \sum_{i=1}^N \hat{\beta}_i^{FC})'. \quad (28c)$$

Let Z_i^p denote the predictive values from the multiequation regression of $Z_i = [Y_i(w) \ X_i]$ on X , the observations for each spatial unit weighted by Φ_i^{-1} ,

$$Z_i^p = X(X\Phi_i^{-1}X')^{-1}X\Phi_i^{-1}Z_i = [Y_i^p(w) \ X_i]. \quad (29)$$

The inverse of Φ_i , Φ_i^{-1} , can be computed by the expression

$$\Phi_i^{-1} = \frac{1}{\sigma_i^2} I_T - \frac{1}{\sigma_i^2} X_i [V^{-1} + \frac{1}{\sigma_i^2} X_i' X_i]^{-1} \frac{1}{\sigma_i^2} X_i', \quad (30)$$

as a result of which the formula for Z_i^p changes into

$$Z_i^p = X \times \left[\frac{1}{\sigma_i^2} X'X - \frac{1}{\sigma_i^2} X'X_i [V^{-1} + \frac{1}{\sigma_i^2} X_i' X_i]^{-1} \frac{1}{\sigma_i^2} X_i' X_i \right]^{-1} \times \left[\frac{1}{\sigma_i^2} X'Z_i - \frac{1}{\sigma_i^2} X'X_i [V^{-1} + \frac{1}{\sigma_i^2} X_i' X_i]^{-1} \frac{1}{\sigma_i^2} X_i' Z_i \right]. \quad (31)$$

Finally, estimate δ_i ($i=1, \dots, N$) and β by

$$\begin{bmatrix} \delta_1 \\ \vdots \\ \delta_N \\ \beta \end{bmatrix} = \begin{bmatrix} Y_1^p(w)' \Phi_1^{-1} Y_1(w) & \cdot & 0 & Y_1^p(w)' \Phi_1^{-1} X_1 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & Y_N^p(w)' \Phi_N^{-1} Y_N(w) & Y_N^p(w)' \Phi_N^{-1} X_N \\ X_1' \Phi_1^{-1} Y_1(w) & \cdot & X_N' \Phi_N^{-1} Y_N(w) & \sum_{i=1}^N X_i' \Phi_i^{-1} X_i \end{bmatrix} \begin{bmatrix} Y_1^p(w)' \Phi_1^{-1} Y_1(w) \\ \cdot \\ Y_N^p(w)' \Phi_N^{-1} Y_N(w) \\ \sum_{i=1}^N X_i' \Phi_i^{-1} Y_i(w) \end{bmatrix}, \quad (32)$$

where each element $M_1' \Phi_i^{-1} M_2$ with $M_1 = Y_1^p(w), \dots, Y_N^p(w), X_1, \dots, X_N$ and $M_2 = Y_1(w), \dots, Y_N(w), X_1, \dots, X_N$ can be computed by

$$M_1' \Phi_i^{-1} M_2 = \frac{1}{\sigma_i^2} M_1' M_2 - \frac{1}{\sigma_i^2} M_1' X_i [V^{-1} + \frac{1}{\sigma_i^2} X_i' X_i]^{-1} \frac{1}{\sigma_i^2} X_i' M_2. \quad (32a)$$

4 CONCLUSIONS

This paper has given a systematic overview of panel data models extended to spatial error autocorrelation or a spatially lagged dependent variable. In the introduction it was stated that these kinds of panel data models are not very well documented in the literature. One reason is that each model has its own specific problems. We summarize these problems for the four panel data models considered in this paper as follows:

- (i) Estimation of the fixed effects model can be carried out with standard techniques developed by Anselin (1988, pp.181-182) and Anselin and Hudak (1995), but the regression equation must first be demeaned. This model is relatively simple.
- (ii) Estimation of the random effects model can be carried out by ML, though two problems must be taken into account. First, estimating the random effects model by ML is not standard in that this application is not available in most commercial econometric software packages. Second, it must be investigated whether the maximum found is global and not local.
- (iii) A heterogeneous model with different coefficients for different spatial units extended to spatial error autocorrelation is equivalent to a seemingly unrelated regressions model. The estimation of this model is standard. A heterogeneous model with different coefficients for different spatial units extended to spatially lagged dependent variables is almost equivalent to a simultaneous linear equations model. Estimation of this model by ML is complicated by the fact that it falls back on the Jacobi iteration method, one that cannot be expected to converge in general. Furthermore, this method is available in only a limited number of commercial econometric software packages. As an alternative, one might use 2SLS, but this method does not take account of the restrictions on the coefficients within the coefficient matrices. Another problem is that the response coefficients in both models cannot be estimated consistently when T would be fixed and $N \rightarrow \infty$, as is typical in short panels. This is because the number of these coefficients increases as N increases.
- (iv) Estimation of the random coefficient model extended to spatial error autocorrelation or to spatially lagged dependent variables by ML is possible though laborious. First, the covariance matrix contains unknown parameters that cannot be solved algebraically from the first-order maximizing conditions of the log-likelihood function. Second, restrictions on these unknown parameters often cause difficulties for existing ML algorithms. Third, existing algorithms that

have solved these problems are usually geared to simpler models without spatial effects. For these reasons it is simpler to use feasible GLS to estimate the random coefficients model extended to spatial error autocorrelation, and to use feasible GLS combined with instrumental variables to estimate the random coefficients model extended to spatially lagged dependent variables. These estimators might still be difficult to compute as they require matrix inversions whose orders may be quite large, dependent on the number of spatial units and the number of explanatory variables. Finally, in the latter model a random element in the coefficients of the spatially lagged dependent variables should be avoided, as this raises intractable difficulties at the levels of both identification and estimation.

REFERENCES

- Anselin L. (1988) Spatial econometrics: methods and models. Kluwer, Dordrecht.
- Anselin L., Hudak S. (1992) Spatial econometrics in practice; a review of software options. *Regional Science and Urban Economics* 22, pp.509-536.
- Balestra P., Negassi S. (1992) A random coefficient simultaneous equation system with an application to direct foreign investment by French firms. *Empirical Economics* 17, pp.205-220.
- Baltagi B.H. (1995) Econometric analysis of panel data. Wiley, Chichester, England.
- Baltagi B.H., Griffin J.M. (1997) Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics* 77, pp.303-327.
- Bowden R.J., Turkington D.A. (1984) Instrumental variables. Cambridge University Press, Cambridge.
- Breusch T.S. (1987) Maximum likelihood estimation of random effects models. *Journal of Econometrics* 36, pp.383-389.
- Buettner T. (1999) The effect of unemployment, aggregate wages, and spatial contiguity on local wages: An investigation with German district level data. *Papers in Regional Science* 78, pp.47-67.
- Case A.C. (1991) Spatial patterns in household demand. *Econometrica* 59, pp.953-965.
- Fortheringham A.S., Charlton M., Brunsdon C. (1997) Measuring spatial variations in relationships with geographically weighted regression. In: Fischer M.M., Getis A. (eds.) Recent developments in spatial analysis, pp. 60-82. Springer, Berlin.
- Goldstein H. (1995) Multilevel statistical models, 2nd edition. Arnold (Oxford University Press), London.
- Greene W.H. (1997) Econometric Analysis, 3rd edition. Macmillan, New York.
- Hausman J.A. (1975) An instrumental variables approach to full information estimators for linear and certain nonlinear econometric models. *Econometrica* 43, pp.727-738.
- Hausman J.A. (1983) Specification and estimation of simultaneous equation models. In: Griliches Z., Intriligator M.D. (eds.) Handbook of econometrics, volume 1, pp.392-448. Elsevier, Amsterdam.
- Hsiao C. (1986) Analysis of panel data. Cambridge University Press, Cambridge.
- Hsiao C. (1996) Random coefficients models. In: Mátyás L., Sevestre P. (eds.) The econometrics of panel data, second revised edition, pp.77-99. Kluwer, Dordrecht.
- Hsiao C., Tahmiscioglu A.K. (1997) A panel analysis of liquidity constraints and firm investment. *Journal of the American Statistical Association* 92, pp.455- 465.
- Jenrich R.I., Schluchter M.D. (1986) Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 42, pp.805-820.
- Kelejian H.H. (1974) Random parameters in simultaneous equations framework; Identification and estimation. *Econometrica* 42: 517-527.
- Kelejian H.H., Robinson D.P. (1997) Infrastructure productivity estimation and its underlying econometric specifications: A sensitivity analysis. *Papers in Regional Science* 76, pp.115-131.
- Kelejian H.H., Prucha I.R. (1999) A generalized moments estimator for the autoregressive

- parameter in a spatial model. *International Economic Review* 40, p.509.
- Lindstrom M.J., Bates D.M. (1988) Newton-Raphson and EM algorithms for linear mixed-effects model for repeated-measures data. *Journal of the American Statistical Association* 83, pp.1014-1022.
- Longford N.T. (1993) Random coefficient models. Clarendon Press, Oxford.
- Maddala G.S. (1971) The use of variance components in pooling cross section and time series data. *Econometrica* 39, pp.341-358.
- Pesaran M.H., Smith R. (1995) Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics* 68, pp.79-113.
- Quah D.T. (1996a) Empirics for economic growth and convergence. *European Economic Review* 40, pp.1353-1375.
- Quah D.T. (1996b) Aggregate and regional disaggregate fluctuations. *Empirical Economics* 21, pp.137-162.
- Sampson R.J., Morenoff J.D., Earles F. (1999) Beyond social capital: spatial dynamics of collective efficacy for children. *American Sociological Review* 64, pp.633-660.
- Shin C., Amemiya Y. (1997) Algorithms for the likelihood-based estimation of the random coefficient model. *Statistics & Probability Letters* 32, pp.189-199.
- Swamy P.A.V.B. (1970) Efficient inference in a random coefficient regression model. *Econometrica* 38, pp.311-323.
- Swamy P.A.V.B. (1974) Linear models with random coefficients. In: Zarembka P. (ed.) *Frontiers in econometrics*, pp.143-168. Academic Press, New York..
- White E.N., Hewings G.J.D. (1982) Space-time employment modeling: Some results using seemingly unrelated regression estimators. *Journal of Regional Science* 22, pp.283-302.